

Highly recommended books:

- **The Cyanobacteria: Molecular Biology, Genomics and Evolution**
- **Microbial Biodegradation: Genomics and Molecular Biology**
- ***Pseudomonas*: Genomics and Molecular Biology**
- ***Acinetobacter* Molecular Microbiology**
- **Bacteriophage: Genetics and Molecular Biology**
- ***Bacillus*: Cellular and Molecular Biology**
- **Coronaviruses: Molecular and Cellular Biology**
- **Real-Time PCR in Microbiology**
- ***Candida*: Comparative and Functional Genomics**

[click here](#)

[click here](#)

[click here](#)

[click here](#)

[click here](#)

[click here](#)

[click here](#)

[click here](#)

[click here](#)

In Vitro Selection of Functional Nucleic Acid Sequences

Dan Nieuwlandt

NeXtar Pharmaceuticals, Inc., Boulder,
Colorado 80301, USA

Abstract

The power of *in vitro* selection methods for the isolation of nucleic acids that display a desired property derives from the enormous number of sequence variants that can be surveyed with relative ease using controlled *in vitro* biochemistry. This methodology has found a variety of applications, ranging from the study of nucleic acid-protein interactions and natural ribozymes to the isolation of nucleic acids with potential as diagnostic or therapeutic reagents or with new catalytic activities. The number of reported applications is growing exponentially, and each application presents new variables and challenges. The goal of this article is to guide prospective users through the myriad decisions that must be made in the design and execution of a successful *in vitro* selection experiment.

Introduction

The SELEX (Systematic Evolution of Ligands by EXponential enrichment) process is a methodology for the *in vitro* selection of nucleic acids capable of a desired function from random sequence libraries (1, 2). SELEX involves reiterative selection and amplification of only those

sequences within a huge population of random sequences (up to 10^{16} different sequences) that display a targeted function. The SELEX process is not only a method for studying the biochemistry of natural nucleic acids, but also a means of isolating nucleic acids with novel properties. Due to intramolecular contacts, including Watson-Crick base pairs, non-standard base pairs, and novel tertiary interactions, a large population of random-sequence single-stranded nucleic acids is synonymous with a large library of stable three-dimensional structures, or shapes, which are dictated by their linear sequences. The overriding premise is that solutions to a wide variety of specific molecular recognition and catalysis problems can be found within such a library. Consistent with this premise, *in vitro* selection experiments have been successfully applied to the selection of nucleic acid ligands (or aptamers; from the Latin *aptus*, to fit; 3) to numerous nucleic acid binding proteins, non-nucleic acid binding proteins, and small organic molecules, as well as to the selection of nucleic acids with catalytic and other properties (reviewed in 2, 4-6).

As illustrated in Figure 1, *in vitro* selection experiments involve the synthesis of a large library of randomized sequences (RNA, ssDNA or dsDNA), the partitioning of functional from non-functional sequences, and amplification of the selected sequences to provide an enriched pool of nucleic acids for the next cycle of selection and amplification. These cycles are typically repeated until those oligonucleotides most capable of the targeted function have evolved from the population. The selected nucleic acids are then isolated by cloning and characterized. While the specific methods employed for

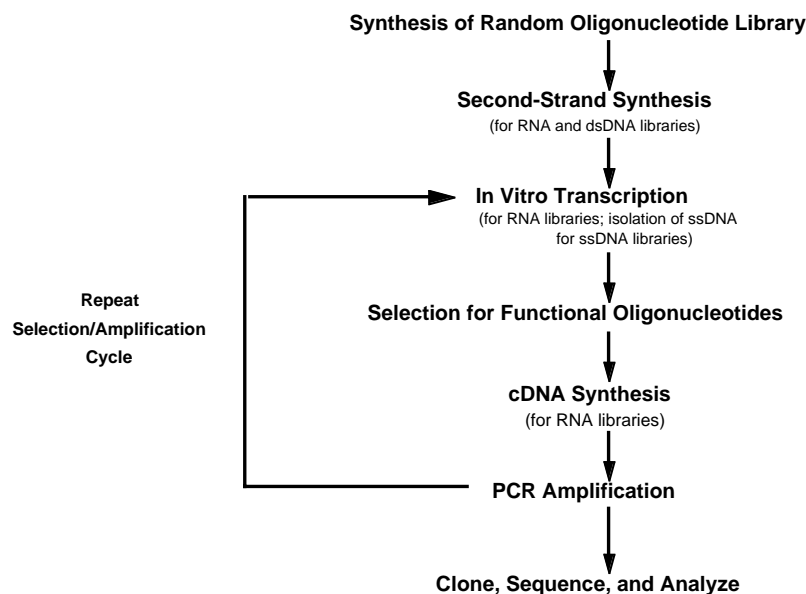


Figure 1. A schematic representation of the SELEX process.

each new application of this process will differ, it is hoped that the options and considerations discussed below will serve as a useful guide.

***In Vitro* Selection Experiments**

An example protocol for a single round of SELEX aimed at the isolation of RNA aptamers, with nitrocellulose filter-binding of RNA-protein complexes as the partitioning method, is provided here as an introduction to the biochemical steps involved in a typical *in vitro* selection experiment and as a basis for the discussion that follows. The selection buffer will be dictated by the target protein and by the desired aptamer properties; RNA and protein concentrations in the binding reaction are adjusted to achieve the desired binding stringency (see text).

Nitrocellulose Filter-binding Selection from an RNA Aptamer Library

(See text for discussion of random library design and construction).

1. In a siliconized eppendorf tube, denature radiolabeled RNA in selection buffer by incubating at 70°C for 5 min, then renature by immediate placement on ice followed by equilibration at 37°C (selection temperature).
2. Add target protein to RNA and incubate at 37°C for 10-20 min. An identical reaction minus the protein should be prepared for background determination (fraction of the RNA pool that is retained on the nitrocellulose filter in the absence of protein).
3. Place nitrocellulose filter (2.5 cm diam., 0.45 µM pore size) on a clean (free of nucleic acids and nucleases) filtration apparatus connected to a vacuum source. Pre-equilibrate the filter by suction-filtration of 5 ml reaction buffer. Suction the binding reaction through the filter and immediately wash with 5 ml of 37°C reaction buffer.
4. Transfer the filter to a clean surface and cut into several pieces with a new scalpel blade. Add the pieces to an eppendorf tube containing 400 µl phenol and 200 µl 7 M urea. Secure lid and incubate with mixing at room temperature for 30 min. Add 100 µl H₂O, vortex-mix, and spin at 14,000 rpm (full speed) in a microcentrifuge for 5 min. Transfer the aqueous phase (top) to a new tube; to the organic phase, add 100 µl H₂O, vortex-mix, and spin as above (back extraction). Combine the two aqueous phases and add 1 µl of 20 µg/µl yeast

tRNA (critical for efficient precipitation of the small quantity of selected RNA), 40 µl of 3 M NaOAc, pH 5.2, and 1 ml of cold 100 % ethanol. Allow to precipitate in a dry ice-ethanol bath for 20 min, then pellet the RNA by centrifugation at 14,000 rpm for 15 min. Carefully remove the supernatant and wash the pellet with cold 80 % ethanol. Resuspend the dried pellet in 20 µl H₂O.

Amplification of Selected RNA

Reverse Transcription

Set up a reverse transcription (RT) reaction as follows:

Selected RNA	20 µl
50 µM 3' primer	2 µl

Incubate RNA + primer at 70°C for 3 min, then cool to room temperature and add:

5X AMV RT buffer	10 µl
10 mM dNTPs	2.5 µl
AMV RT	20 U
H ₂ O	to 50 µl

Incubate reaction at 45°C for 45 min. 5X AMV RT buffer contains: 250 mM Tris-HCl, pH 8.3 at 45°C, 300 mM NaCl, 30 mM Mg(OAc)₂, and 50 mM DTT. Recover cDNA by ethanol precipitation and suspend in 40 µl H₂O.

PCR Amplification

1. Set up a pilot PCR amplification to determine the optimal number of temperature cycles. Set up the reaction using 25 % of the recovered cDNAs:

cDNA	10 µl
10X <i>Taq</i> buffer	10 µl
25 mM MgCl ₂	15 µl
10 mM dNTPs	6 µl
50 µM 5' primer	4 µl
50 µM 3' primer	4 µl
<i>Taq</i> polymerase	5 U
H ₂ O	to 100 µl

All reaction components except polymerase are combined and incubated at 94°C for 3 min. Add polymerase immediately following incubation and overlay the reaction mixture with 50 µl mineral oil (keep reaction at or above annealing temperature). Amplify using the temperature cycle parameters listed below (for the primers listed in Figure 2). 10X *Taq* buffer: 500 mM KCl, 100 mM Tris-HCl, pH 9.0, 1 % Triton X-100. Magnesium, dNTP, primer, and enzyme concentrations

Synthetic ssDNA:	5'-GGGAGACAAGAATAAACGCTCAA-[N] _n -TTCGACAGGAGGCTACAACAGGC-3'
5' primer:	5'- <u>TAATACGACTCACTATAGGG</u> GAGACAAGAATAAACGCTCAA-3'
3' primer:	5'-GCCTGTTGTGAGCCTCCTGTCGAA-3'
5' cloning primer:	5'-CAG <u>AAGCTT</u> AATACGACTCACTATAGGGAGACAAGAATAAACGCTCAA-3'
3' cloning primer:	5'-GACTGGATCCGCCTGTTGTGAGCCTCCTGTCGAA-3'

Figure 2. Example of an oligonucleotide set suitable for the *in vitro* selection of RNA sequences. The T7 promoter sequence within the 5' primer, and *Hind* III and *Bam*HI restriction endonuclease recognition sites within the 5' and 3' cloning primers, respectively, are underlined.

should be optimized with the appropriate primers and cDNA template prior to initiation of a SELEX experiment.

Temperature cycles:	
94°C	45 sec
58°C	45 sec
72°C	1 min

Remove 3 μ l samples of the reaction after, for example, 12, 14, 16, 18, and 20 cycles. By analysis of the samples on a non-denaturing polyacrylamide gel (alongside the appropriate dsDNA size markers), determine the optimal number of temperature cycles for amplification.

- Set up three reactions identical to that above and PCR amplify with the optimal number of temperature cycles. Pool products (including remainder of pilot reaction) and recover the DNA from the aqueous phase by ethanol precipitation. Suspend the dried pellet in 100 μ l H₂O.

Transcription

Set up a T7 RNA polymerase transcription reaction as follows:

DNA template	100 pmoles (archive remainder at -70°C)
5X T7 buffer	80 μ l
10 mM rNTPs	40 μ l
[α - ³² P]ATP	5 μ l (800 Ci/mmol, 10 μ Ci/ μ l)
T7 RNA polymerase	40 μ g
H ₂ O	to 400 μ l

Incubate at 37°C for 3-6 hours. Add 20 U DNase I and continue incubation at 37°C for 30 min. 5X T7 buffer: 200 mM Tris-HCl, pH 8.0, 60 mM MgCl₂, 30 % PEG8000, 25 mM DTT, 0.01 % Triton X-100. Purify full-length transcripts on a denaturing polyacrylamide gel. Recover RNA from gel slices by electroelution or by passive elution into 300 mM NaOAc, pH 5.2, 2 mM EDTA. Pass eluted RNA through a cellulose acetate membrane (0.2-0.45 μ m pore size). Recover RNA by ethanol precipitation (with at least two 80 % ethanol washes) or, preferably, by concentration on a 10-30,000 MWCO filter (e.g., microcon microconcentrators; Amicon, Inc., Beverly, MA) followed by two 400- μ l H₂O washes. Suspend RNA in H₂O and quantitate. The enriched pool of RNA is now ready for the next cycle of selection and amplification.

Random Sequence Nucleic Acid Libraries

All nucleic acid libraries utilized for *in vitro* selection experiments are composed of a randomized region flanked by defined sequences that permit amplification (and for RNA selections, transcription). The length of the randomized region and the degree of randomization are largely dictated by knowledge of the target and the purpose of the experiment. Strategies have included patch randomization, partial (or degenerate) randomization, and complete randomization. Patch randomization involves the randomization of a short block, or blocks, of sequence within a nucleic acid of known sequence and function for the purpose of determining the optimal sequence(s) for the substructure under study. In the first SELEX experiment, Tuerk and Gold (1) took this approach by randomizing only

the eight-base loop region of a 36-nt stem-loop RNA structure that interacts with T4 DNA polymerase. This permitted an exhaustive sampling of each of the 65,536 (4⁸) possible sequences within this critical region of the RNA structure.

Partial randomization of a known protein-binding or catalytic nucleic acid, such that the sequence is biased towards the wild-type by only a partial randomization at each nucleotide position (e.g., 80% wild-type, 20% non-wild-type nucleotide at each position), may be a useful approach if the goal is to optimize or alter the properties of the existing sequence and/or to decipher the higher-order structure of the nucleic acid. Since each selected functional nucleic acid would be related, a comparative analysis (7) of the sequence variants should prove useful for establishing structural features important for function. The extent of randomization would depend on the length of the sequence required for function; in general, a lower level of randomization should be used with longer sequences if a significant number of functional sequence variants are to remain in the population. Bartel *et al.*, (8) have used 30% randomization at each nucleotide position to study the structure of the Rev-binding site within a 66-nucleotide segment of HIV-I mRNA (about 20 nucleotides were required for function), while 5% randomization at each position has been used to study a 127-nt *sunY* ribozyme variant (9).

Most *in vitro* selection experiments have been performed without any preconceptions about what structures might emerge from the selection and have therefore been initiated with oligonucleotide libraries containing sequences with a completely randomized block of contiguous nucleotides. The major issue in this case is the length of the randomized sequence. Contiguous random sequence lengths of 30 to 120 nucleotides have typically been used, and nucleic acid aptamers with high affinity and specificity for their targets have been isolated consistently with libraries at the lower end of this range (2). If the sampling of all possible sequences of a given length is desired, then the realistic upper limit on this length is about 25 nucleotides. The number of oligonucleotide molecules that gives a 99% likelihood of complete sequence representation is: $(2.30 \times L + 7.65) \times 10^{-24} \times 4^L$ moles; where L = random sequence length (10). For oligonucleotides with 25 completely randomized positions, 7.34×10^{-8} moles (73.4 nmoles; 4.42×10^{16} molecules) would be required for this degree of sequence representation. For a 72-nt oligonucleotide (25 random positions plus the defined primer binding sequences shown in Figure 2), 1.74 mg would be required. The required quantity becomes prohibitive for a 26-nt randomized sequence (304 nmoles; 1.83×10^{17} molecules; 7.22 mg).

While stable structures can certainly be formed with 25 nucleotides, it is expected that the chances of isolating an oligonucleotide with the desired properties would be improved with longer random regions. Longer sequences would have the advantage of a greater number of registers of a given short sequence, resulting in increased representation of that sequence and theoretically an increase in the representation of short structural motifs, presented in various contexts, that can be assumed by that sequence. More important, many functions, such as high affinity and specificity for certain targets and the catalysis of certain reactions, may be achieved only with

complex structures requiring longer sequences. Although the 4^{100} possible sequences 100-nt in length, for example, would be only minutely sampled with a nucleic acid library consisting of 10^{15} - 10^{16} oligonucleotides, sequences that might assume a "scaffold" structure required for proper positioning of critical short invariant sequences may have only loose primary sequence requirements. The disadvantages of using longer randomized sequences include the increased difficulty in chemically synthesizing the initial ssDNA pool (discussed below) and in analyzing the selected sequences. If the purpose of an *in vitro* selection is to obtain an oligonucleotide-based reagent that will be needed in large quantities, the use of shorter sequences in the experiment will ensure that selected sequences can be efficiently chemically synthesized. In addition, with an increase in sequence length comes an increase in the number of energetically stable, and potentially inactive, alternative secondary structures that can be formed by an individual sequence. While this could present a serious problem if a structurally homogenous preparation of a selected sequence is required, it can be argued that, given a strong selective advantage, only those sequences which can consistently fold into the active conformation would be isolated.

Synthesis of Oligonucleotide Libraries

The starting point for SELEX experiments, whether the library is to be composed of ssDNA, dsDNA, or RNA, is the generation of a pool of oligodeoxynucleotides with a standard DNA synthesizer. Mixtures of phosphoramidites are prepared for coupling at positions that are to be randomized. An equimolar mixture of the four phosphoramidites usually leads to a bias towards the coupling of G and T; for positions that are to be completely randomized, using a mixture with an A:C:G:T ratio of 3:3:2:2 should provide a more even distribution of the four bases at each position (11). Equally important considerations are differences in the shelf-lives of the reagents and minor variations in the flow rate between reagent lines. For these reasons, fresh phosphoramidites should be prepared and the appropriate mixture of these should be added to a bottle that is attached to one of the spare ports on the machine. The machine can be programmed to draw from this bottle for random couplings. For partially randomized positions, the mixtures are simply adjusted to give the desired ratio of wild-type to non-wild-type bases. Sequence bias in the randomized region of the oligonucleotide product can be detected by base composition analysis (12). The synthesized oligonucleotides are enzymatically degraded to their constituent nucleosides with snake venom phosphodiesterase and bacterial alkaline phosphatase. The nucleosides are then separated by reversed-phase (C18) HPLC. A comparison of the U.V. absorbance data to that obtained with authentic nucleoside standards permits accurate quantitation. Alternatively, significant sequence bias can be detected by sequencing a sample of the synthesized pool using standard dideoxynucleotide termination methods. Although less accurate than the HPLC method, randomized regions where each of the four bases are close to being equally represented should result in equal band intensities across all four gel lanes (the defined primer binding sequence serves as an internal control for the efficiency of incorporation of the four

dideoxynucleotides). A combinatorial approach, where oligonucleotides from separate syntheses are joined, must be taken to construct oligonucleotide libraries when the desired sequence length exceeds the useful limits of DNA synthesis (120-150 nucleotides, depending on stepwise coupling yields). One approach is to convert the synthetic oligos into dsDNA fragments that, by virtue of restriction enzyme sites, can be ligated together (13). An alternative method involves synthesizing the oligonucleotides with 8-10 nucleotides of complementarity between their 3' termini. The annealed oligonucleotides can be extended with Klenow enzyme to generate the full length dsDNA template (i.e., primer extension). Both of these methods result in a random region which is interrupted by a short (6-10 nucleotides) fixed sequence.

The defined primer-binding sequences flanking the randomized region should be chosen with potential PCR artifacts in mind. These sequences should be long enough to permit high specificity during amplification (20-24 nt), and, to avoid self-priming and primer-dimer artifacts, the two primers should not be highly structured or share significant regions of complementarity. In addition, if several selection experiments are being performed in the same laboratory, it is a good idea to use different primer binding sequences for each library to prevent cross-contamination. For RNA selections, one of the primer sequences must include an RNA polymerase promoter. Although many RNA polymerases are available, T7 RNA polymerase, which is relatively inexpensive and is the most amenable to large scale transcription reactions, is the enzyme of choice (14). For long synthetic ssDNA pool templates, the promoter sequence is often omitted to reduce the length of the oligonucleotide and thus increase synthesis yields; the promoter is introduced by one of the primers during PCR amplification. To facilitate cloning at the completion of the selection experiment, a separate set of primers with restriction enzyme recognition sites flanking the primer sequences used during the selection process are often used. A few extra bases flanking the restriction sites may be required for efficient digestion. An example of an oligonucleotide set suitable for an RNA SELEX experiment is provided in Figure 2.

The purified synthetic ssDNA must be converted to double-stranded DNA (dsDNA) to generate a template pool for *in vitro* transcription (for RNA experiments) or to generate the starting library for dsDNA selections. Using the oligonucleotides in Figure 2 as an example, large scale (or numerous small scale) two-cycle PCR reactions, with 1 μ M template DNA, 1.5 μ M of each primer, and prolonged extension times (10-15 min) are sufficient to copy the synthetic strand and to introduce the strand primed by the oligonucleotide containing the T7 promoter. The pool complexity can be determined by gel-purifying and quantitating the larger promoter-containing template (since each full-length enzymatically-synthesized template should correspond to a single oligonucleotide from the original synthetic pool). A determination of pool complexity by this or a similar method becomes particularly important when large synthetic oligonucleotides (> 100 nucleotides) are used; a significant fraction of the synthetic DNA can suffer from synthesis-induced lesions that prevent second strand synthesis.

For both RNA and ssDNA selections, a renaturation protocol is introduced prior to contact with the target during

each selection cycle. The purpose is to allow the oligonucleotides to dissociate from energetically stable alternative conformations and intermolecular contacts so that they can assume their most stable conformations in the chosen selection buffer. A common protocol involves heating the sample at an elevated temperature (70-90°C, 3-5 minutes), followed by either slow-cooling to the selection temperature or immediate placement on ice followed by equilibration at the selection temperature. Which protocol is followed is not as important as following the identical denaturation/renaturation protocol prior to each selection cycle.

Selection Methods

Any selection method that effectively partitions the desired functional oligonucleotides from inactive (or weakly functional) nucleic acids is potentially useful in an *in vitro* selection experiment. Methods that have been successfully exploited to selectively retain oligonucleotide-target complexes include nitrocellulose filter-binding, targets immobilized on solid supports (e.g., agarose, polymethyl methacrylate, or paramagnetic beads), and gel mobility shift. As described above, filter-binding selections are performed by suction filtering the oligonucleotide/protein mixture through a nitrocellulose filter, followed by removal of remaining unbound oligonucleotides by washing with the selection buffer. Oligonucleotide-protein complexes retained on the filter are eluted and amplified (1). Oligonucleotide ligands can similarly be retained with bead-coupled targets (using either column or batch methods). Following extensive washing to remove unbound oligonucleotides, the retained species can be recovered by denaturing the oligonucleotide-target complex, by affinity elution, or by cleavage of the target from the solid support.

While denaturing (dissociating) the oligonucleotide-target complex (with EDTA, urea, or phenol, for example) will effectively elute target-bound species, it is risky because oligonucleotides bound to matrix components other than the target will also be eluted. When a sufficient quantity of free target is available, affinity elution will virtually assure that ligands specific for the free target are isolated (15). A disadvantage of affinity elution is that ligands with faster off-rates will be preferentially eluted; the concentration of free target and time of incubation are therefore important considerations. The use of solid supports that contain a cleavable linker between the target and the matrix are an alternative to affinity elution (16). Target-bound oligonucleotides would be selectively eluted by cleavage of the linker. Gel-mobility shift partitioning is most appropriate for the selection of dsDNA-protein interactions (17). High background problems will often be encountered with single-stranded nucleic acids due to intermolecular interactions and the aberrant migration of various oligonucleotide structures (which may co-migrate with target-bound oligonucleotides).

Complex Targets

Although purified targets have been utilized for most SELEX experiments, complex targets such as whole cells, membrane preparations, and protein mixtures are amenable to this technique as well. Many methods employed for receptor binding studies (18) would be

suitable partitioning methods for whole-cell or membrane SELEX experiments, while immunoprecipitation is an option for recovering ligands to specific targets within protein mixtures (19). Individual oligonucleotides isolated from a complex target SELEX may serve as a reagent for the identification of their specific targets.

Catalytic Oligonucleotides

Oligonucleotides with catalytic activity can be effectively partitioned from inactive oligonucleotides if this activity results in a selectable modification of the catalyst. For many ribozyme selections, this modification has been the conversion of a non-amplifiable template to an amplifiable template. Bartel and Szostak (13, 20) selected RNAs that catalyze 3'-5' phosphodiester bond formation between the catalyst (RNA ligase) and a substrate RNA sequence. A primer complementary to the ligated substrate sequence was utilized for both affinity column selection and PCR amplification. Pan and Uhlenbeck (21) selected RNAs that undergo autocatalytic cleavage with Pb^{2+} ; circularized RNAs that were linearized (cleaved) in the presence of the metal were selectively ligated to primer annealing sequences and amplified. Other selections have used substrates tagged with a selectable group. Ribozymes that catalyze reactions resulting in covalent attachment of the substrate can be partitioned by virtue of the tag. Wilson and Szostak (22) isolated self-alkylating ribozymes that catalyze carbon-nitrogen bond formation between the catalyst and a substrate tagged with biotin. Biotinylated RNAs were selected with bead-coupled streptavidin. Using a similar partitioning scheme, Lohse and Szostak (23) isolated ribozymes with acyl transferase activity by selecting for the transfer of a biotinylated methionine from an oligonucleotide substrate to the catalyst. RNAs that catalyze the transfer of the γ -thiophosphate of ATP- γ S to the 5'-hydroxyl or internal 2'-hydroxyls of the catalyst have also been isolated (24). In this case, the ribozymes were partitioned by chromatography on a thiopyridine-activated thiopropyl-Sepharose column (disulfide bond formation between the matrix and RNAs containing thiophosphate groups). An approach that has been investigated with only limited success involves the selection of oligonucleotides that bind to a transition state analogue, followed by assaying the selected ligands for their ability to catalyze the targeted reaction (2, 25, 26).

Reducing Background

The potential for retention of non-specific nucleic acids (background) exists for any partitioning method, with the effect of increasing the number of rounds required to isolate the "winners" (or failure of the experiment). It is important to recognize the additional "targets" present in the selection reaction and to have a means of removing sequences that bind to these (a process that has been termed counterSELEX, or negative selection). Filter-specific binders, for example, can often be removed by first passing the oligonucleotide pool through a series of filters prior to addition of the target. Oligonucleotides that bind to components of a solid support (other than the target) can be removed by first exposing the pool to the identical matrix minus the target, followed by addition of the unbound oligonucleotides to the target-coupled support. Nucleic acid

aggregates can be a problem for many partitioning protocols. Large aggregates can be removed by centrifugation for about 10 minutes at full speed in a microcentrifuge; size-exclusion chromatography may be a useful alternative approach. A powerful method for eliminating background problems is to alternate between different partitioning protocols throughout the *in vitro* selection experiment. Non-specific oligonucleotides retained on agarose beads, for example, may be eliminated by partitioning with paramagnetic beads or nitrocellulose filters in subsequent selection cycles.

Selection Stringency

The first two or three SELEX rounds, when individual sequences are not highly represented, should be performed under low stringency conditions to ensure the recovery of rare functional sequences. Once a high copy number of the functional sequences is obtained, selection stringency can be increased to demand competition between the different species capable of binding to the target. It can be assumed that weakly functional sequences far outnumber highly functional sequences, so multiple rounds with progressively increasing stringency will be required to enrich the pool for the best sequences. Increasing binding reaction stringency generally means reducing the target concentration while keeping the oligonucleotide:target ratio at 10-100:1. An important consideration is the fraction of the pool that partitions in the absence of target (background; see above section). The fraction of the library that specifically partitions with the target should be significantly greater (preferably at least 10-fold greater) than the background fraction before stringency is increased to prevent background oligonucleotides from overwhelming the pool population. Irvine *et al.*, (10) have provided a thorough analysis of selection stringency factors. An *in vitro* selection is usually considered to have been completed when the activity of the oligonucleotide pool plateaus (i.e., a further increase in stringency does not lead to an increase in pool activity).

Amplification Methods

Reverse transcription with AMV reverse transcriptase (for RNA selections) and PCR amplification with *Taq* DNA polymerase have been the most common methods for amplifying selected oligonucleotides. An isothermal transcription-based amplification system (TAS) has also been used for RNA selections (27, 28). With TAS, the selected RNAs, reverse transcriptase, T7 RNA polymerase, dNTPs, rNTPs, and primers are combined in a single reaction mixture. RNA is copied to cDNA by reverse transcriptase (with RNase H activity) and the resulting cDNA is transcribed by T7 RNA polymerase (100-1,000 copies per cDNA template), yielding an amplification of up to 10⁶-fold after 1 hour at 37°C. The overriding concern when deciding upon an amplification protocol should be reaction efficiency. Both reverse transcription and PCR amplification reaction conditions (or TAS reaction conditions) should be optimized with the appropriate RNA and cDNA template preparations prior to the start of a new SELEX experiment. High efficiency amplification will prevent a potentially disastrous loss of functional sequences during the critical first selection/amplification

cycle (when copy number may be as low as one) and will keep amplification biases to a minimum.

For *in vitro* selection experiments with ssDNA, a method of differentiating the two strands is required to facilitate purification of the sense strand following PCR amplification. This can be accomplished by a variety of methods, including asymmetric PCR (29), incorporating a cleavable bond into one of the primers (30), or, most commonly, by biotinylating one of the primers. In this latter case, biotinylated PCR products can be bound to bead-coupled streptavidin followed by a denaturing step to recover the non-biotinylated strand (31). The use of a primer with three biotins coupled to the 5' end, introduced during DNA synthesis with biotin phosphoramidites (available from Glen Research, Sterling, VA, <http://www.glenred.com/>; Cat. No. 10-191953-xx), improves the efficiency of this process and allows an alternative protocol: biotinylated strands up to at least 100-nt in length (the longest attempted in our laboratory) can be separated from non-biotinylated strands by standard denaturing PAGE. Electrophoresis on a sequencing-length 6% denaturing gel until the xylene cyanol marker dye has migrated about 30 cm is sufficient for the separation of biotinylated 100-nt ssDNAs (retarded migration) from their complementary non-biotinylated ssDNAs.

Analysis of Selected Sequences

During and after an *in vitro* selection experiment, the base distribution at each nucleotide position can be estimated by sequencing a sample of the pool by standard reverse transcriptase (for RNA pools) or DNA polymerase dideoxy sequencing methods. This is particularly useful for patch randomization selections where the evolution of preferred sequences can be visualized by comparing sequencing results from sequential selection cycles (1).

Individual members of the selected pool are isolated by cloning and sequenced. An alignment and comparative analysis of sequences isolated from patch or partial randomization studies should be straightforward (since all sequences should be related), but this can often be a challenge when selections are performed with completely random libraries. To facilitate the analysis, sequences are usually assigned to classes on the basis of sequence similarities. The ability of representatives of each class to perform the targeted function is verified, followed by an alignment of the sequences within each functional class. "Information boundary" experiments are usually performed at this stage to determine the 5' and 3' boundaries of the sequence required for function (32). A comparative sequence analysis will generally provide reliable structural information. With this process, the alignment is examined for sites within putative structures where base-pairing is maintained despite nucleotide substitutions (covariation). The structure is likely to exist if multiple instances of covariation are observed among the sequences under study (7, 33). Energy minimization programs (34) provide thermodynamically optimal and suboptimal secondary structures, which can be of help (but can also be misleading) by pointing out potential structures that can be examined by comparative sequence analysis criteria. An energy minimization package (mfold version 2.3) can be accessed through the internet (<http://www.ibc.wustl.edu>). Useful complementary data can also

be obtained with enzymatic (35) and chemical (36) structure probing.

Increasing Sequence Diversity with PCR

As mentioned above, the structural solutions to many functions may not be found in a sequence length that permits all possible sequences of that length to be sampled (~25 nucleotides). While functional oligonucleotides can be isolated from libraries composed of long completely randomized regions, where the sequence space is only sparsely sampled, the chances that the optimal sequence for a particular function will be isolated is low. However, by introducing additional sequence diversity during the *in vitro* selection process, the sequence space will be more thoroughly sampled and sequences more fit for the targeted function should evolve from the oligonucleotide population (see Chapters 3, 10-12). This *in vitro* evolution process occurs to some extent during all *in vitro* selection experiments due to the use of polymerases that lack proofreading activities. As demonstrated by Beaudry and Joyce (37), the mutagenesis rate can be accelerated by carrying out the PCR amplification step under conditions that decrease the fidelity of *Taq* DNA polymerase (13, 16, 28, 38). Error-prone PCR, as this has been termed, can of course also be applied to any gene of interest (including sequences obtained through the SELEX process) for the purpose of obtaining sequences with increased fitness, or for obtaining a family of functional sequence variants for comparative sequence analysis. The effect of reaction conditions on *Taq* DNA polymerase fidelity is discussed in detail in Chapter 3 and in the references cited above.

The SELEX process is being aggressively applied to the discovery of new drugs and diagnostic reagents and to basic research problems. Still, the utility of SELEX as a research tool has only begun to be exploited. Our understanding of nucleic acid structure and function, and of the catalytic potential of nucleic acids, should advance at a rapid pace with increasing applications of this methodology.

Acknowledgements

I thank Dr. Larry Gold for critical review and helpful comments and my colleagues at NeXstar Pharmaceuticals, Inc. and the University of Colorado for stimulating discussions.

References

1. Tuerk, C. and Gold, L. 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*. 249: 505-510.
2. Gold, L., Polisky, B., Uhlenbeck, O. and Yarus, M. 1995. Diversity of oligonucleotide functions. *Annu. Rev. Biochem.* 64: 763-797.
3. Ellington, A.D. and Szostak, J.W. 1990. *In vitro* selection of RNA molecules that bind specific ligands. *Nature*. 346: 818-822.
4. Gold, L. 1995. Oligonucleotides as research, diagnostic, and therapeutic agents. *J. Biol. Chem.* 270: 13581-13584.
5. Conrad, R.C., Giver, L., Tian, Y. and Ellington, A.D. 1995. *In vitro* selection methodologies to probe RNA function and structure. *Mol. Div.* 1: 69-78.
6. Uphoff, K.W., Bell, S.D. and Ellington, A.D. 1996. *In vitro* selection of aptamers: The dearth of pure reason. *Curr. Opin. Structural Biol.* 6: 281-288.
7. Woese, C.R. and Pace, N.R. 1993. Probing RNA structure, function, and history by comparative analysis. In: *The RNA World*. R.F. Gesteland and J.F. Atkins, eds. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York. p. 91-117.
8. Bartel, D.P., Zapp, M.L., Green, M.R. and Szostak, J.W. 1991. HIV-I Rev regulation involves recognition of non-Watson-Crick base pairs in viral RNA. *Cell*. 67: 529-536.
9. Green, R. and Szostak, J.W. 1992. Selection of a ribozyme that functions as a superior template in a self-copying reaction. *Science*. 258: 1910-1915.
10. Irvine, D., Tuerk, C. and Gold, L. 1991. SELEXION: Systematic Evolution of ligands by exponential enrichment with integrated optimization by non-linear analysis. *J. Mol. Biol.* 222: 739-761.
11. Lato, S.M., Boles, A.R. and Ellington, A.D. 1995. *In vitro* selection of RNA lectins: Using combinatorial chemistry to interpret ribozyme evolution. *Chem. Biol.* 2: 291-303.
12. Eadie, J.S., McBride, L.J., Efcavitch, J.W., Hoff, L.B. and Cathcart, R. 1987. High-performance liquid chromatographic analysis of oligodeoxynucleotide base composition. *Anal. Biochem.* 165: 442-447.
13. Bartel, D.P. and Szostak, J.W. 1993. Isolation of new ribozymes from a large pool of random sequences. *Science*. 261: 1411-1418.
14. Milligan, J.F. and Uhlenbeck, O.C. 1989. Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol.* 180: 51-62.
15. Jenison, R.D., Gill, S.C., Pardi, A. and Polisky, B. 1994. High resolution molecular discrimination by RNA. *Science*. 263: 1425-1428.
16. Nieuwlandt, D., Wecker, M. and Gold, L. 1995. *In vitro* selection of RNA ligands to substance P. *Biochem.* 34: 5651-5659.
17. He, Y.-Y., Stockley, P.G. and Gold, L. 1996. *In vitro* evolution of the DNA binding sites of *Escherichia coli* methionine repressor, MetJ. *J. Mol. Biol.* 255: 55-66.
18. Hulme, E. C. 1992. Receptor-ligand interactions: a practical approach. Oxford University Press, New York.
19. Tsai, D.E., Harper, D.S. and Keene, J.D. 1991. U1-snRNP-A protein selects a ten nucleotide consensus sequence from a degenerate RNA pool presented in various structural contexts. *Nucleic Acids Res.* 19: 4931-4936.
20. Eklund, E.H., Szostak, J.W. and Bartel, D. P. 1995. Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science*. 269: 364-370.
21. Pan, T. and Uhlenbeck, O.C. 1992. *In vitro* selection of RNAs that undergo autocatalytic cleavage with Pb²⁺. *Biochem.* 31: 3887-3895.
22. Wilson, C. and Szostak, J.W. 1995. *In vitro* selection of a self-alkylating ribozyme. *Nature*. 374: 777-782.
23. Lohse, P.A. and Szostak, J.W. 1996. Ribozyme-catalyzed amino-acid transfer reactions. *Nature*. 381: 442-444.

24. Lorsch, J.R. and Szostak, J.W. 1994. *In vitro* evolution of new ribozymes with polynucleotide kinase activity. *Nature*. 371: 31-36.
25. Prudent, J.R., Uno, T. and Schultz, P.G. 1994. Expanding the scope of RNA catalysis. *Science*. 264: 1924-1927.
26. Morris, K.N., Tarasow, T.M., Julin, C.M., Simons, S.L., Hilvert, D. and Gold, L. 1994. Enrichment for RNA molecules that bind a Diels-Alder transition state analog. *Proc. Natl. Acad. Sci. USA*. 91: 13028-13032.
27. Guatelli, J.C., Whitfield, K.M., Kwoh, D.Y., Barringer, K.J., Richman, D.D. and Gingeras, T.R. 1990. Isothermal, *in vitro* amplification of nucleic acids by a multienzyme reaction modeled after retroviral replication. *Proc. Natl. Acad. Sci. USA*. 87: 1874-1878.
28. Lehman, N. and Joyce, G.F. 1993. Evolution *in vitro* of an RNA enzyme with altered metal dependence. *Nature*. 361: 182-185.
29. Dicker, A.P., Volkenandt, M. and Bertino, J.R. 1993. Manual and direct sequencing of product generated by the polymerase chain reaction. In: *Methods in Molecular Biology*, Vol. 15: PCR Protocols: Current Methods and Applications. B. A. White, ed. Humana Press, Totowa, NJ. p.143-152.
30. Silveira, M.H. and Orgel, L.E. 1995. PCR with detachable primers. *Nucleic Acids Res.* 23: 1083-1084.
31. Bock, L.C., Griffin, L.C., Latham, J.A., Vermass, E.H. and Toole, J.J. 1992. Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature* 355: 564-566.
32. Jellinek, D., Green, L.S., Bell, C. and Janjic, N. 1994. Inhibition of receptor binding by high-affinity RNA ligands to vascular endothelial growth factor. *Biochem.* 33: 10450-10456.
33. James, B.D., Olsen, G.J. and Pace N.R. 1989. Phlogenetic comparative analysis of RNA secondary structure. *Methods Enzymol.* 180: 227-239
34. Zuker, M. 1989. Computer prediction of RNA structure. *Methods Enzymol.* 180: 262.
35. Knapp, G. 1989. Enzymatic approaches to probing of RNA secondary and tertiary structure. *Methods Enzymol.* 180: 192-227.
36. Ehresmann, C., Baudin, F., Mougel, M., Romby, P., Ebel, J.-P. and Ehresmann, B. 1987. Probing the structure of RNAs in solution. *Nucleic Acids Res.* 15: 9109-9128.
37. Beaudry, A.A. and Joyce, G.F. 1992. Directed evolution of an RNA enzyme. *Science*. 257: 635-641.
38. Cadwell, R.C. and Joyce, G.F. 1992. Randomization of genes by PCR mutagenesis. *PCR Methods Appl.* 2: 28-33.